DSSTox Log File:
NCTR Estrogen Receptor Binding Database (NCTRER)
(last updated 15 February 2008)

Description: Information in this file documents the creation, review, and update process for the DSSTox NCTRER SDF file and provides summary information on database content. The first section summarizes the process used for creating the initial DSSTox SDF files, and the quality assurance checks and procedures employed. A table providing field and data counts offers summary overview of NCTRER database content. The Log Table documents modifications and revisions to the database content or format in version updates. To obtain the most current version of this Log File and a record of any new modifications, or to report errors in this file, a user should consult the DSSTox NCTRER database page:
http://www.epa.gov/ncct/dsstox/sdf_nctrer.html

QA and Development Notes for v1a:
The original NCTR ER database was provided by the NCTR Source (Weida Tong) in SDF format. This file contained basic chemical information (Structure, CASRN, Chemical Name) and ER relative binding activity information. We thank Weida Tong and Hong Fang for invaluable assistance in various stages of development and quality review, providing additional data fields (**LogP**), clarifying numerous issues pertaining to the experimental data, and approving all added field names and contents. Chemical structures provided by the NCTR Source in the original SDF were rendered as three dimensional (Corina, version 2.3). For construction of the main DSSTox SDF, these structures were redrawn in more easily viewed 2D form using CambridgeSoft ChemFinder (ver 7.0 for Windows). We consulted the main citation [Fang et al., 2001, Chem. Res. Tox., 14:280-294] and confirmed structures and CAS numbers at the ChemFinder website (http://chemfinder.cambridgesoft.com/). CambridgeSoft ChemFinder (ver 7.0 for Windows) was also used for automatic generation of SMILES codes from structures. There are only two chemicals classified as "organometallic" and 4 "defined organics" classified as either salt or complex in the NCTRER.

Notes for v2a:
For version 2a, a variety of fields have been added. IUPAC systematic chemical names, **ChemName_IUPAC**, were computed by Marc Nicklaus (NCI) using the ACD Labs IUPAC Name-Generation software (ACD/NameBatch, version 8.05). **INChI** codes were automatically generated from the final DSSTox SDF using a pre-release version of the publicly available program, wINChI11b.exe, accessible from the NIST INChI developers (http://chemdata.nist.gov/IChI/INChIv11b.zip).

Notes for v3a,b:
Revised DSSTox Standard Chemical Fields are included (see http://www.epa.gov/ncct/dsstox/MoreonStandardChemFields.html) along with updated InChI codes (version 1.0), recomputed IUPAC chemical names (ACDLabs ACD/Name, version 9.0), and many regenerated 2D structures with stereochemistry of steroidal compounds rendered in more standardized form. Additionally, an extensive quality review of all DSSTox chemical records was performed, resulting in numerous corrections and modifications to chemical structures and added information (CASRN, representative structures for mixtures, etc) throughout DSSTox data files. For more information on current review procedures, see
http://www.epa.gov/ncct/dsstox/ChemicalInfQAProcedures.html

Notes for v4a:
NCTRER_v4 has no new chemical records but has several minor QA corrections, field entry revisions, field changes, new CASRN, etc. Changes to DSSTox Standard Chemical Fields include new ID fields: **DSSTox_RID**, **DSSTox_Generic_SID** and **DSSTox_FileID** (replacing **DSSTox_SID** and **DSSTox_ID_FileName**) (see http://www.epa.gov/ncct/dsstox/MoreonStandardChemFields.html). NCTRER-specific chemical information has been

removed from the **ChemicalNote** field and moved to new Source-specific field, **Note_NCTRER**. Also, entries in **TestSubstance_Description** field have been simplified. Finally, text entries in field **Mean_ER_RBA_ChemClass** have been convert it to a pure numeric field.

Notes for v4b:
NCTRER_v4b includes minor structure changes/modifications (stereochemistry) and two new summary activity fields for use in PubChem and structure-activity relationship studies: **ActivityOutcome_NCTRER** (entries of active, inactive, or inconclusive) and **ActivityScore_NCTRER** (INTEGER[0-100]). In addition, the new **STRUCTURE_InChIKey** field (25 character abbreviated InChI for use in structure-indexing applications) has been added as a DSSTox Standard Chemical Field to all DSSTox files

Log of SDF Modifications and Version/revision updates:

Date	DSSTox SDF File Name	Modifications from previous version	Additional Notes
23Oct03	NCTRER_v1a_232_23Oct03 NCTRER_DOP_v1a_230_23Oct03 NCTRER_DOP3D_v1a_230_23Oct03	Initial launch publication; no previous published versions.	NCTRER is considered a "static" historical database meaning that further expansion of the database to include additional data is unlikely. Future updates will correct reported errors provided by users or incorporate DSSTox format changes.
1Mar05	NCTRER_v2a_232_1Mar05	New Fields: **INChI**, **ChemName_IUPAC**, **StudyType**, **Species**, **Endpoint** Addition of **SMILES_Parent** to Main file. Added CAS for DSSTox_ID=43, 136. Removed incorrect CAS for 36,37 (CAS for parent not derivative in both cases).	Major format modification to include INChI, IUPAC names, and ToxML fields. Separate "desalted" defined organic parent (DOP) file not provided. Users can easily generate DOP file by extracting "defined organic" records and converting **SMILES_Parent** to structures.
2May05	NCTRER_v2a_232_2May05_ nostructures.xls	User reported duplicate field columns unintentionally included in file; corrected.	No other NCTRER files modified at that time.

10Apr2006	NCTRER_v3b_232_10Apr2006	Updated with new DSSTox Standard Chemical Fields and entries (*revised Aug 2005*).
		Field **Rationale ChemClass ERB** renamed **ActivityCategory_Rational_ChemClass_ERB** to be more consistent with usage in other DSSTox files.
		Renamed Source-related fields to eliminate spaces:
		   ChemClass_ERB
		   LOG_ER_RBA
		   ActivityCategory_ER_RBA
		   Mean_ER_RBA_ChemClass
		   F1_Ring
		   F2_AromaticRing
			Numerous structure modifications and changes in stereochemical rendering throughout DSSTox data files following major quality review.
			
			Thanks to Andreas Karwath for ErrorReport.
			
			NCTRER_v3a _232_22Oct2005:
			Note: earlier version of this file was provided to PubChem, with identical format to v3b but latter has undergone additional QA review and has small number of corrections/modifications.

		F3_PhenolicRing **F4_Heteroatom** **F5_Phenol3nPhenyl** **F6_OtherKeyFeatures** Updated InChI codes (version 1.0). Updated IUPAC chemical names (ACDLabs Name to Structure, version 9.0). Expanded "ddmmmyear" format for dates in DSSTox file names (e.g., 26Aug2005). Corrected F5_Phenol3nPhenyl field contents for DSSTox_ID=14, 28, 40, 41, 42 (changed entry from 1 to 0) based on ErrorReport.	

15Jun2007	NCTRER_v4a_232_15Jun2007	Revised Standard Fields:
		   DSSTox_SID has been replaced by two new ID fields **DSSTox_RID** and **DSSTox_Generic_SID**.
		   DSSTox_ID_FileName has been replaced by new ID field: **DSSTox_FileID**.
		   Entries in **TestSubstance_Description** field have been simplified.
		   Entries in **ChemicalNote** that pertained specifically to NCTRER have been moved to Source-Specific field: **Note_NCTRER**.
		
		The field, **Mean_ER_RBA_ChemClass**, was stripped of text and converted to pure numeric entries.
		Endpoint field entry changed from "ER RBA" to "Estrogen Receptor Relative Binding Affinity"	

15Feb2008	NCTRER_v4b_232_15Feb2008	Total of 11 modifications to structures, adding or modifying stereochemistry:
		
		New Standard Field added:
		   STRUCTURE_InChIKey
		
		Two new summary activity fields added in coordination with PubChem deposits:
		   ActivityOutcome_NCTRER
		   ActivityScore_NCTRER
			All corrections or changes to structure information noted in **Note_NCTRER** field, searchable by version (e.g., v4b).

Field and Data Counts in Older Version DSSTox SDF files:

DSSTox SDF	Standard Chemical Fields	Standard Toxicity Fields	Source-specific fields	Chemical records total	STRUCTURE_ChemicalType:		
					Defined Organic	Inorganic	Organo-metallic
NCTRER_v1a	14	0	13	232*			
NCTRER_DOP_v1a	16	0	13	230*			
NCTRER_DOP3D_v1a	16	0	13	230*			
NCTRER_v2a	17	3	13	232*	230	0	2

* NCTRER contains 6 sets of alpha, beta stereochemical pairs that have the same 2D representation but different 3D structures and different activities.

NCTRER SDF Content*	Totals_v3b	Totals_v4a	Totals_v4b
# Records	232	232	232
DSSTox Standard Chemical Fields	18	18	19
DSSTox Standard Toxicity Fields	3	3	3
NCTRER Source Fields	13	14	16
Total # Fields	34	35	38
Chemical Content	**Counts_v3b**	**Counts_v4a**	**Counts_v4b**
STRUCTURE_ChemicalType:			
defined organic	230	230	230
inorganic	0	0	0
organometallic	2	2	2
no structure	0	0	0
STRUCTURE_TestedForm_ DefinedOrganic:			
parent	226	226	226
complex	3	3	3
salt	1	1	1
salt complex	0	0	0
TestSubstance_Description:			
single chemical compound	225	225	225
defined mixture or formulation	7	* (NA)	* (NA)
undefined mixture	0	* (NA)	* (NA)
macromolecule	0	0	0
mixture or formulation	* (NA)	7	7

* (NA) = field entry not applicable for DSSTox file version indicated

Chemical Counts and Activity Distribution in Main Structural Classes of NCTRER:

ChemClass_ERB	Active Strong	Active Medium	Active Weak	Slight Binder	Inactive	Total # of Chemicals	Mean_ER_RBA _ChemClass
Steroids	12	7	3	0	9	31	1.24
DES	13	6	2	1	0	22	2.14
Phytoestrogens	3	13	18	0	12	46	0.019
Diphenylmethanes	1	6	11	3	9	30	0.0087
Biphenyls	0	2	6	1	3	12	0.0028
Phenols	0	4	19	1	5	29	0.00088
Misc	0	3	2	2	55	62	NA
Totals	29	41	61	8	93	232	
Summed Totals	Actives		131	Inactives	101	232	

Wanted!! CASRN Information

The listing below provides chemicals with known structures and Unknown CASRN entries, which is primarily an indication of the little studied nature of these particular chemicals in the CPDB. For each, a CAS registry search was performed in CAS SciFinder and no CASRN was found by the CPDB Source authors. However, if a user has new information pertaining to any Unknown CASRN in the below listing, please report this using a DSSTox Error Report Form that can be accessed from any DSSTox SDF Download Page, and be sure to indicate all relevant information (**DSSTox_RID**, **TestSubstance_ChemicalName**, nature of missing information, source of correct information, etc.). Thank you!

DSSTox_ RID	TestSubstance_ChemicalName	STRUCTURE_SMILES	CASRN	Date of Request
22343	6-hydroxy-2'-methoxy flavone	C1(=CC(=O)C2C(O1)=CC=C(O)C=2)C3=C(OC)C=CC=C3	Unknown	15Jun2007
22344	3'-hydroxy flavanone	O1C(CC(=O)C2C1=CC=CC=2)C3C=C(O)C=CC=3	Unknown	15Jun2007
22501	meso-p-(alpha,beta-diethyl-p-methyl-phenethyl)-phenol	C(C(C1=CC=C(C=C1)C)CC)(C2C=CC(=CC=2)O)CC	Unknown	15Jun2007
22504	2,6-dimethyl hexestrol	C1(C(C(C2=CC=C(O)C=C2)CC)CC)=CC(C)=C(C(=C1)C)O	Unknown	15Jun2007
22539	3,6,4'-trihydroxyflavone	C1(=C(OC2=CC=C(C=C(C1=O)2)O)C3=CC=C(C=C3)O)O	Unknown	15Jun2007